# NVM Express Technical Errata

| | |
|---|---|
| **Errata ID** | **011** |
| **Change Date** | **2/6/2014** |
| **Affected Spec Ver.** | **NVM Express 1.0 and NVM Express 1.1a** |
| **Corrected Spec Ver.** | |

Submission info

| Name | Company | Date |
|---|---|---|
| Amber Huffman | Intel | 1/15/2014 |
| Judy Brock | Samsung | 1/15/2014 |
| Ken Okin | HGST | 1/15/2014 |
| Brad Besmer | LSI | 1/16/2014 |
| Matthew Lovell | Seagate | 2/5/2014 |
| Peter Onufryk | PMC-Sierra | 2/5/2014 |

The erratum updates the Media Errors status codes to be clear that it also includes Data Integrity type errors.

A clarification is added for the CC.Enable field that namespaces may not be ready when CC.EN is set to '1'.

A clarification is added that IO commands may fail while a format is in progress.

A few clarifications for metadata and protection information have been added.

A clarification was made to not require the PCI Power Management capability to be the first capability.

A clarification for the Do Not Retry bit in Status for cases of successful completion was added.

The erratum clarifies interrupt masking in relation to the MSI capability structure.

This erratum clarifies that Critical Warning bits represent the current state and are not persistent.

Description of the specification technical flaw:

***Modify Figure 29 as shown below:***

**Figure 29: Status Code – Status Code Type Values**

| Value | Description |
|---|---|
| 0h | **Generic Command Status:** Indicates that the command specified by the Command and Submission Queue identifiers in the completion queue entry has completed. These status values are generic across all command types, and include such conditions as success, opcode not supported, and invalid field. |
| 1h | **Command Specific Status:** Indicates a status value that is specific to a particular command opcode. These values may indicate additional processing is required. Status values such as invalid firmware image or exceeded maximum number of queues is reported with this type. |
| 2h | **Media and Data Integrity Errors:** Any media specific errors that occur in the NVM or data integrity type errors shall be of this type. |
| 3h – 6h | Reserved |
| 7h | **Vendor Specific** |

***Modify section 4.6.1.2.3 as shown below:***

### 4.6.1.2.3 Media and Data Integrity Errors Definition

Completion queue entries with a Status Code Type of Media and Data Integrity Errors indicate an error associated with the command that is due to an error associated with the NVM media or a data integrity type error.

**Figure 34: Status Code – Media and Data Integrity Error~~s~~ Values**

| Value | Description |
|---|---|
| 00h – 7Fh | Reserved |
| 80h – BFh | I/O Command Set Specific |
| C0h – FFh | Vendor Specific |

**Figure 35: Status Code – Media and Data Integrity Error~~s~~ Values, NVM Command Set**

| Value | Description |
|---|---|
| 80h | **Write Fault:** The write data could not be committed to the media. ~~This may be due to the lack of available spare locations that may be reported as an asynchronous event.~~ |
| 81h | **Unrecovered Read Error:** The read data could not be recovered from the media. |
| 82h | **End-to-end Guard Check Error:** The command was aborted due to an end-to-end guard check failure. |
| 83h | **End-to-end Application Tag Check Error:** The command was aborted due to an end-to-end application tag check failure. |
| 84h | **End-to-end Reference Tag Check Error:** The command was aborted due to an end-to-end reference tag check failure. |
| 85h | **Compare Failure:** The command failed due to a miscompare during a Compare command. |
| 86h | **Access Denied:** Access to the namespace and/or LBA range is denied due to lack of access rights. Refer to TCG SIIS. |
| 87h – BFh | Reserved |

***Modify bytes 175:160 in Figure 75 as shown below:***

| | |
|---|---|
| 175:160 | **Media and Data Integrity Errors:** Contains the number of occurrences where the controller detected an unrecovered data integrity error. Errors such as uncorrectable ECC, CRC checksum failure, or LBA tag mismatch are included in this field. |

***Modify the Enable (EN) field in section 3.1.5 as shown below:***

| | | | |
|---|---|---|---|
| 00 | RW | 0 | **Enable (EN):** When set to '1', then the controller shall process commands based on Submission Queue Tail doorbell writes. When cleared to '0', then the controller shall not process commands nor post completion queue entries to Completion Queues. When this field transitions from '1' to '0', the controller is reset (referred to as a Controller Reset). The reset deletes all I/O Submission Queues and I/O Completion Queues, resets the Admin Submission Queue and Completion Queue, and brings the hardware to an idle state. The reset does not affect PCI Express registers nor the Admin Queue registers (AQA, ASQ, or ACQ). All other controller registers defined in this section and internal controller state (e.g., Feature values defined in section 5.12.1 that are not persistent across power states) are reset to their default values. The controller shall ensure that there is no data loss for commands that have had corresponding completion queue entries posted to an I/O Completion Queue prior to the reset operation. Refer to section 7.3 for reset details.<br><br>When this field is cleared to '0', the CSTS.RDY bit is cleared to '0' by the controller once the controller is ready to be re-enabled. When this field is set to '1', the controller sets CSTS.RDY to '1' when it is ready to process commands. CSTS.RDY may be set to '1' before namespace(s) are ready to be accessed.<br>< INSERT BLANK LINE><br>Setting this field from a '0' to a '1' when CSTS.RDY is a '1,' or setting this field from a '1' to a '0' when CSTS.RDY is a '0,' has undefined results. The Admin Queue registers (AQA, ASQ, and ACQ) shall only be modified when EN is cleared to '0'. |

***Modify the third paragraph of section 5.13 as shown below:***

The Format NVM command shall fail if the controller is in an invalid security state. See the TCG SIIS reference. The Format NVM command may fail if there are outstanding IO commands to the namespace specified to be formatted. IO commands for a namespace that has a Format NVM command in progress may fail.

***Modify byte 26 in Figure 84 as shown below:***

| | | |
|---|---|---|
| 26 | M | **Formatted LBA Size (FLBAS):** This field indicates the LBA data size & metadata size combination that the namespace has been formatted with (refer to section 5.13).<br><br>Bits 7:5 are reserved.<br><br>Bit 4 if set to '1' indicates that the metadata is transferred at the end of the data LBA, creating an extended data LBA. Bit 4 if cleared to '0' indicates that all of the metadata for a command is transferred as a separate contiguous buffer of data.<br><br>Bits 3:0 indicates one of the 16 supported combinations indicated in this data structure. This is a 0's based value. |

***Modify byte 08 in Figure 111 as shown below:***

| | |
|---|---|
| 08 | **Protection Information Location (PIL):** If set to '1' and protection information is enabled, then protection information is transferred as the first eight bytes of metadata. If cleared to '0' and protection information is enabled, then protection information is transferred as the last eight bytes of metadata. This setting is reported in the Formatted LBA Size field of the Identify Namespace data structure. |

***Modify the first paragraph of section 5.12.1.8 as shown below:***

This Feature configures interrupt coalescing settings. The controller should signal an interrupt when either the Aggregation Time or the Aggregation Threshold conditions are met. If either the Aggregation Time or the Aggregation Threshold fields are cleared to 0h, then an interrupt may be generated (i.e., interrupt coalescing is implicitly disabled). This Feature applies only to the I/O Queues. It is recommended that interrupts for commands that complete in error are not coalesced. The settings are specified in Command Dword 11.

***Modify Figure 100 as shown below:***

**Figure 100: Interrupt Coalescing – Command Dword 11**

| Bit | Description |
|---|---|
| 31:16 | Reserved |
| 15:08 | **Aggregation Time (TIME):** Specifies the recommended maximum time in 100 microsecond increments that a controller may delay an interrupt due to interrupt coalescing. A value of 0h corresponds to no delay ~~(i.e., disabling this capability)~~. The controller may apply this time per interrupt vector or across all interrupt vectors. The reset value of this setting is 0h. |
| 07:00 | **Aggregation Threshold (THR):** Specifies the recommended ~~desired~~ minimum number of completion queue entries to aggregate per interrupt vector before signaling an interrupt to the host. This is a 0's based value. The reset value of this setting is 0h. |

***Modify the first paragraph of section 5.10.1.2 as follows:***

This log page is used to provide SMART and general health information. The information provided is over the life of the controller and is retained across power cycles. The log page shall be supported on a global basis. To request the global log page, the namespace specified is FFFFFFFFh. The log page may also be supported on a per namespace basis, as indicated in the Identify Controller data structure in Figure 82. If the log page is not supported on a per namespace basis, specifying any namespace other than FFFFFFFFh should abort the command with status Invalid Field in Command. There is not namespace specific information defined in the SMART / Health log page in this revision, thus the global log page and namespaces specific log page contain identical information.

***Modify the Capabilities Pointer in section 2.1.19 as shown below:***

### 2.1.19 Offset 34h: CAP – Capabilities Pointer

| Bit | Type | Reset | Description |
|---|---|---|---|
| 7:0 | RO | Impl Spec ~~PMCAP~~ | **Capability Pointer (CP):** Indicates the first capability pointer offset. ~~It points to the PCI Power Management capability offset.~~ |

***Modify Figure 28 as shown below:***

**Figure 1: Completion Queue Entry: Status Field**

| Bit | Description |
|---|---|
| 31 | **Do Not Retry (DNR):** If set to '1', indicates that if the same command is re-submitted it is expected to fail. If cleared to '0', indicates that the same command may succeed if retried. If a command is aborted due to time limited error recovery (refer to section 5.12.1.5), this field should be cleared to '0'. If the SCT and SC fields are cleared to 0h then this field should be cleared to '0'. |
| 30 | **More (M):** If set to '1', there is more status information for this command as part of the Error Information log that may be retrieved with the Get Log Page command. If cleared to '0', there is no additional status information for this command. Refer to section 5.10.1.1. |
| 29:28 | Reserved |
| 27:25 | **Status Code Type (SCT):** Indicates the status code type of the completion queue entry. This indicates the type of status the controller is returning. |
| 24:17 | **Status Code (SC):** Indicates a status code identifying any error or status information for the command indicated. |

***Modify the INTMS register in section 3.1.3 as shown below:***

| Bit | Type | Reset | Description |
|---|---|---|---|
| 31:00 | RW1S | 0h | **Interrupt Vector Mask Set (IVMS):** This field is bit significant. If a '1' is written to a bit, then the corresponding interrupt vector is masked from generating an interrupt or reporting a pending interrupt in the MSI Capability Structure. Writing a '0' to a bit has no effect. When read, this field returns the current interrupt mask value within the controller (not the value of this register). If a bit has a value of a '1', then the corresponding interrupt vector is masked. If a bit has a value of '0', then the corresponding interrupt vector is not masked. |

***Modify the text of section 7.5.1 as shown below:***

This is the mode of interrupt operation if any of the following conditions are met:

- Pin based interrupts are being used – MSI (MSICAP.MC.MSIE='0') and MSI-X are disabled
- Single MSI is being used – MSI is enabled (MSICAP.MC.MSIE='1'), MSICAP.MC.MME=0h, and MSI-X is disabled
- Multiple MSI is being used – Multiple-message MSI is enabled (MSICAP.MC.MSIE='1') and (MSICAP.MC.MME=1h) and MSI-X is disabled.

Within the controller there is an interrupt status register (IS) that is not visible to the host. In this mode, the IS register determines whether the PCI interrupt line shall be driven active or an MSI message shall be sent. Each bit in the IS register corresponds to an interrupt vector. The IS bit is set to '1' when the AND of the following conditions is true:

- There is one or more unacknowledged completion queue entries in a Completion Queue that utilizes this interrupt vector;
- The Completion Queue(s) with unacknowledged completion queue entries has interrupts enabled in the "Create I/O Completion Queue" command;
- The corresponding INTM bit exposed to the host is cleared to '0', indicating that the interrupt is not masked.

For single and multiple MSI, the INTM register masks interrupt delivery prior to MSI logic. As such, an interrupt on a vector masked by INTM does not cause the corresponding Pending bit to assert within the MSI Capability Structure.

If MSIs are not enabled, IS[0] being a one causes the PCI interrupt line to be active (electrical '0'). If MSIs are enabled, any change to the IS register that causes an unmasked status bit to transition from zero to one or clearing of a mask bit whose corresponding status bit is set shall cause an MSI to be sent. Therefore, while in

wire mode, a single wire remains active, while in MSI mode, several messages may be sent, as each edge triggered event on a port shall cause a new message.

In order to clear an interrupt for a particular interrupt vector, host software shall acknowledge all completion queue entries for Completion Queues associated with the interrupt vector.

***Modify byte 0 in Figure 75 as shown below:***

<table>
<tr>
<td>0</td>
<td>Critical Warning: This field indicates critical warnings for the state of the controller. Each bit corresponds to a critical warning type; multiple bits may be set. If a bit is cleared to ‘0’, then that critical warning does not apply. Critical warnings may result in an asynchronous event notification to the host. Bits in this field represent the current associated state and are not persistent.
<table>
<tr><th>Bit</th><th>Definition</th></tr>
<tr><td>00</td><td>If set to ‘1’, then the available spare space has fallen below the threshold.</td></tr>
<tr><td>01</td><td>If set to ‘1’, then the temperature has exceeded a critical threshold.</td></tr>
<tr><td>02</td><td>If set to ‘1’, then the device reliability has been degraded due to significant media related errors or any internal error that degrades device reliability.</td></tr>
<tr><td>03</td><td>If set to ‘1’, then the media has been placed in read only mode.</td></tr>
<tr><td>04</td><td>If set to ‘1’, then the volatile memory backup device has failed. This field is only valid if the controller has a volatile memory backup solution.</td></tr>
<tr><td>07:05</td><td>Reserved</td></tr>
</table>
</td>
</tr>
</table>

Disposition log

| | |
|---|---|
| 1/15/2014 | Erratum captured. |
| 1/29/2014 | Added interrupt coalescing clarifications, added SMART / Health log page global versus namespace specific clarification, modified write error handling language, clarified that PCI Power Management is not required to be the first capability. |
| 2/5/2014 | Added clarification for Do Not Retry bit, added interrupt clarifications, added Critical Warning clarification. |
| 2/6/2014 | Removed write error handling to be covered in ECN 012. |
| 3/31/2014 | Erratum ratified. |